Humanistic Mathematics Network Journal

Issue 8 Article 22

7-1-1993

On Learning in the Mathematical Sciences: Statistics 200 as a Paradigm of Everything Wrong in Mathematical Education

James M. Cargal
Troy State University, Montgomery

On Learning in the Mathematical Sciences: Statistics 200 as a Paradigm of Everything Wrong in Mathematics Education

James M. Cargal
Troy State University in Montgomery
Montgomery, AL
36103

One of the hazards in mathematics teaching is that it is possible to communicate without teaching. The litmus test of learning should be whether the student knows anything six months after the course is over. Too often, there is no intellectual evidence that the student ever took the course; and too often it is the fault of the course. This hazard can exist in any mathematical course at any level (and with *good* students) but a superb example is the elementary course in statistics, which I will call *Statistics 200*. The distilled thesis of this essay is that many courses try to teach too much material too fast. To illustrate this point, it is useful to dissect a particular course and that course will be Statistics 200. Note that this essay is intended partly as a sequel to, but is independent of [1] where I tried to address the problem of communication in the classroom. Note, that I agree with most statisticians that statistics, unlike probability, is outside of mathematics proper; nonetheless it is a mathematical discipline and statistical courses suffer most of the problems of mathematics courses.

Too often, there is no intellectual evidence that the student ever took the course; and too often it is the fault of the course.

Virtually every university and college in the United States has a sophomore introductory course in statistics. This course is usually required of business majors and social science majors and is frequently taken by *hard* science majors who may later take courses in mathematical statistics. A legitimate question is whether any learning at all ever takes place in the course. *Does as many as one student in fifty leave the course with any reward commensurate to the effort expended?*

It can be argued that the above question is inherently unfair. How often is the efficacy of any specific course tested post-course? GRE's and SAT's and that ilk of tests target broad areas. But individual courses are not subjected to this sort of analysis. Nor are instructor evaluations given post-course. It would be interesting to acquire evaluations of the teacher and the course (say) six months after a course; time can dramatically change the focus of perceptions. No more will be said here about teachers; our concern is with curriculum design, and in that area we do test the student's post-course knowledge of the material whenever we give a course that has the other as a prerequisite. Note, that teachers nearly always feel that their students have inadequate knowledge of the prerequisites!

The Statistics 200 course is used here as a paradigm of what goes wrong in mathematics teaching. I am not interested in the problems of teaching that course per se, but since I am using it, I will have something to say about teaching statistics and probability.

Statistics 200

The primary problem with teaching Statistics 200 is that the course has a huge mass of ideas that students with little background are expected to master. They are expected to learn the rudiments of probability theory. Then they are expected to learn data collection and data analysis through hypothesis testing, linear regression, and analysis of variance. *Many* teachers expect more: for example, nonparametric tests and Bayesian methods. The only prerequisite for the course is

having passed algebra. As a rule none of the students will have any prior knowledge of probability or statistics. Furthermore, those students who are taking the course as a requirement for a degree in psychology or business, usually have not had algebra for a while, and are weak in that area too. My experience at several institutions and the experience of others I have talked to, is that even the engineering and science majors retain little of the course, despite the fact many of them do not have a difficult time with it.

The Problem of Probability

Usually Statistics 200 starts with two or three weeks devoted to probability. The idea is that the student needs probability to understand the concepts of statistics. Let us look at a few of the laws of probability that we try to impart in this short time:

The law of addition: $P(A \text{ or } B) = P(A + B) = P(A) + P(B) - P(A \text{ and } B)$

The definition of conditional probability:

$$P(A \text{ given } B) = P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

The law of multiplication:

$$P(A \text{ and } B) = P(AB) = P(B)\cdot P(A|B) = P(A)\cdot P(B|A)$$

The definition of independence: Events A and B are independent if any of the following *equivalent* conditions are true

$$P(A) = P(A|B)$$
$$P(B) = P(B|A)$$
$$P(AB) = P(A)\cdot P(B)$$

Bayes' rule: Given that the events B_i are exhaustive and exclusive

$$P(B_j|A) = \frac{P(A|B_j)\cdot P(B_j)}{\sum_{i=1}^{n} P(A|B_i)\cdot P(B_i)}$$

Notice that I have not said anything about sample spaces and events. I always introduce these things informally by writing out the full 36-point sample space of the outcome of throwing a pair of ordinary die. I then use events within this sample space to illustrate all the rules above except Bayes' rule (which I usually do not cover). I generally define the following events:

S_i is the event that the sum of the dice is i, i = 2,3,4,...,12.

R_i is the event that the red die is i, i = 1,2,3,4,5,6.

These two sets of events are sufficient for illustrating all the rules above.

Many teachers give so much attention to the addition rule for disjoint events that the students do not realize it is a special case of the general rule above, and they do not realize that the above rule is *always* true. At this stage, most students struggle

---

GRE's and SAT's and that ilk of tests target broad areas. But individual courses are not subjected to this sort of analysis. Nor are instructor evaluations given post-course.

---

to differentiate between *or's* and *and's* and have a major difficulty distinguishing between *A and B* and *A given B*. The simplest type of environment to give problems on this is the probability matrix such as:

	C	D	E
A	2	2	1
B	2	3	0

In this setting each event is either a row or a column. The numerical entries are precisely the probabilities of joint events. To solve for P(A|D) we can use the formula directly to get:

$P(A|D) = \frac{P(AD)}{P(D)} = \frac{.2}{.5} = .4$. Or we can legitimately resort to arm waving: the event A makes up .2 out of the event D which has a probability of .5 and that gives A a probability of .2 out of .5, or just .4. Not only do probability matrices give the simplest sort of problems, but

they tend to demonstrate the difficulty the students have with the concepts. No matter how simple such concepts seem to the teacher, they are difficult to digest for the students.

In the probability matrix above, the only events that are independent are A and C, and B and C. I stress independence is a matter of *information:* two events are independent if the occurrence of one event does not effect the probability of the other event, that is it yields no information about the other. This in fact is a literal restatement of parts two and three of the definition of independence. I find that I have to stress repeatedly that the three statements are equivalent and I have to stress precisely what that means. Events A and B above are dependent precisely because they are disjoint, if one occurs, then the other can not occur. We can verify this by recourse to the definition. Yet, no matter how many times one repeats this type of example, many students will cling to the idea of disjoint events being independent events. To them, *disjoint* seems to be *independent.* An example that I have found to get a lot of student response is by using the experiment above of throwing a pair of fair dice. The example is that the events S7 and R4 are independent but the event S6 and R4 are dependent. I have found this example to generate a lot of interest and questions.

A further example that I like can be found in Hamming [2 p. 23]. The problem of the gold coins is as follows: There are three drawers each containing two coins. One contains two gold coins; one contains two silver coins; and one contains a gold and a silver coin. Having picked a

There are so many ideas inherent in the above material that by covering it all in a short time, the student can wind up absorbing none of it.

drawer and a coin at random (with equal probability) and found the coin to be gold, what is the probability that the other coin in that drawer is gold? What is wonderful about this problem is that it is counterintuitive and can be solved by a simple application of the definition of condition probability. Let G_i and S_i ($i = 1,2$) be the events that the first or second coin is gold or silver. Our question is to find P(G2|G1). By definition:

$$P(G_2 \backslash G_1) = \frac{P(G_1G_2)}{P(G_1)} = \frac{\frac{1}{3}}{\frac{1}{2}}$$

Given the difficulty of basic concepts in probability, I find it hard to understand why a beginning class (at the sophomore level) would be given Bayes' rule. It is irrelevant to basic statistics and it is notationally and conceptually more difficult than the concepts discussed so far. I have in the past taught decision theory to students that did not have a background in probability and were not particularly mathematical. In that case the difficulty is the same, but Bayes' rule is necessary. Rather than formally stating Bayes' rule, I have taught them to solve Bayesian problems using probability trees. I will not go into this further since it is not relevant to my main thrust. Bayesian problems can be rendered so simple through probability trees that virtually any student can master the technique.

Discrete probability frequently comes down to a matter of problems in counting. Needless to say, few of the students will have any background here. One topic that we might find useful here is the binomial coefficient. It lends itself to many counting problems such as probabilities in poker. Since it is also required for the binomial distribution, we might as well introduce it during discrete probability and just before the binomial distribution. Normally we define the binomial coefficient *n choose m* as $\binom{n}{m} = \frac{n!}{m!(n-m)!}$ This definition is easy to motivate by beginning with the number of permutations of m out of n items. It needs to be stressed that the above definition is not a good computational definition. We normally calculate binomial coefficients by the usual device of judicious cancelling. This is in fact roughly the way we would do the computation by computer. To compute n choose m, we replace m by n - m if n - m is smaller than m. We then use: $\binom{n}{0} = 1$; *else* $\binom{n}{m} = \frac{n}{m}\binom{n-1}{m-1}$ We are now prepared to go into the binomial and related distributions.

The Central Limit Theorem

The Central Limit Theorem (CLT) is not only one of the most elegant and surprising results in probability theory, but it is the foundation of much of statistics. As a result, many textbooks and teachers give CLT extensive attention. Furthermore there are physical as well as computer aids to demonstrate CLT. I myself used to use spreadsheets. I would have a spreadsheet column containing 200 entries each of which would be a sum of uniform (pseudo) random numbers. I would use the spreadsheet's built in frequency tool and graphing capabilities to graph the frequency distribution of the 200 sums. This would be approximately normally distributed.

It is not surprising that many instructors will devote an entire lecture to CLT. It is a subject many of us love, and any good instructor is likely to emphasize its beauty and importance. Here we touch upon the core problem with Stat 200, and the theme of this essay. A month after the course is over, how many of the students can state or paraphrase CLT? How many even remember it, if reminded of its content? If you say 10% or more, you are either a truly great teacher or you have great students.

Further Comments on Probability

If I have spent too much time on probability, consider that in Statistics 200 we are expected to give all of this information in three weeks or less (and there are topics I left out). There are so many ideas inherent in the above material that by covering it all in a short time, the student can wind up absorbing none of it. For example most people take a while to learn to calculate probabilities with binomial coefficients. Understanding the theory is one thing but there is an art to calculating discrete probabilities. In the recent novel *House of Cards* by Conall Ryan [3] (who is himself a computer scientist) the main character calculates that there are 41 hands that will beat four Kings in five-card poker. We can count these ourselves: there are forty straight-flushes and one hand of four aces. However, there are in actuality forty-eight hands of four aces, reflecting the forty-eight possibilities for the fifth card. Later he calculates that there are forty-three hands that will beat four jacks: the same mistake. In fact this is still not quite correct since we have not considered how the cards showing effect the counting.

The Vos Savant Affair

Discrete probability is somewhat less theoretical than continuous probability but can be quite a bit trickier. Any person who puts toc much faith in their probabilistic intuition is probably a fool. There is an abundance of counter-intuitive probability problems that stump nearly everyone the first time they see them. Such a problem is the Vos Savant teaser. Marilyn Vos Savant (who is reportedly listed in the book of Guiness records as having the world's highest I.Q.) has a column that appears in Parade magazine that comes with many Sunday papers. In one issue [4] a problem occurred that achieved a great deal of controversy and led to subsequent columns and then attention in academic journals [5]. The problem is, like all good counter intuitive teasers, quite simple. A game show consists of choosing which of three curtains hides a great prize as opposed to the other two curtains which will conceal turkey prizes. The contestant having chosen one curtain, the game show host reveals one of the other two curtains to conceal nothing and then offers the contestant the chance to choose the remaining curtain. The question is, should the contestant switch? Vos Savant gave the correct answer, which is yes. Typical reasoning says there is no advantage in switching, that there are two remaining curtains and the chances are fifty-fifty. Vos Savant's reasoning is that the initial probabilities were one-third to two-thirds and that nothing has changed. The host used his knowledge to pick a curtain that would be empty and the probabilities are

With time, the substance of any course that we master seems to diminish into a nice tidy package. It seems reasonable that we can teach the material, to capable students, in very short time, and when the students start to choke on the material it is easy to conclude that the difficulty is that the students are not capable.

unchanged, that is, the originally chosen curtain has probability of one-third to contain the great prize, and the remaining curtain has probability of two-thirds. The problem also lends itself to simple simulation which can even be carried out as a mental experiment. (This is exactly how Hamming solves his problem I quoted above.)

After publishing the problem and solution, Vos Savant received a quantity of condemning letters from scientists many of whom felt it necessary to reveal that they had Ph.D.'s (and therefore couldn't be wrong) and that she was apparently stupid to make such an idiotic mistake.

I find this whole episode to be disturbing. The Vos Savant problem is not a particularly counter-intuitive problem and it has one element that should given any problem solver pause; the game show host possesses and apparently uses information, specifically he knows which curtain has the prize and which is empty.

Another disturbing point is the number of *authorities* who were arrogant and the number who even failed to consider the problem worth thought. Whether you agree with Vos Savant or not, the problem always deserved thought. Her solution can still be quibbled with, for example see [5] or [6]. I myself am disturbed by the apparent manner that so-called experts present themselves to others. I think they reflect badly on the entire mathematical community. They hold up their Ph.D.'s as talismans: *I have great magic and the mathematical spirit speaks through me and if you disagree with me you are damned.*

Probability is clearly not an easy subject to assimilate for anyone. Yet in Statistics 200 we expect students who have little training and who often are not mathematically inclined to absorb the great body of the subject in three weeks.

Statistics

So far we have looked at the probability that usually comprises the first two to four weeks of Stat 200. Again, I am not saying that it *should* comprise the beginning of the course. I think maybe we should consider an elementary course in probability (and probabilistic reasoning) as a prerequisite for the elementary course in statistics. But for the time being most of us are constrained by our current curriculum to provide both topics in one course.

Our first problem is that the students generally have no idea what statistics is. They are taking the course because it is required. Oddly enough we tend to consider a statistics course successful if and only if the students become somewhat adept at applying statistical formulae. However, this ability does not imply any real appreciation of statistic's role in society and in science. Furthermore such an appreciation can be given without learning any formulas. A book that does this well is *Statistics: A guide to the Unknown* [7]. I believe that this book, or some equivalent book, should be required in every introductory course on statistics.

Now, I give you the prime exhibit of this essay, its raison d'etre. We are going to motivate the formula for a confidence interval. To make things easier we will assume a sample from a normal population with known standard deviation; hence we make no appeal to the Central Limit Theorem nor do we have to use the t distribution. The population consists of independently distributed observations from a normal population with known mean and unknown standard deviation s. Our sample is randomly drawn and of size **n**. We use the fact that the sample means (for samples of size n) are also normally distributed with the same mean and with standard deviation $\frac{s}{\sqrt{n}}$. Using the fact that a normal distribution contains 95% of its population within 1.96 standard deviations of the mean we have:

$$P\left(m-1.96\frac{s}{\sqrt{n}} < \bar{x} < m + 1.96\frac{s}{\sqrt{n}}\right) = .95$$

. Subtracting m from all three terms inside the parentheses we get:

$$P\left(-1.96\frac{s}{\sqrt{n}} < \bar{x} - m < +1.96\frac{s}{\sqrt{n}}\right) = .95$$

.

Now by subtracting $\bar{x}$ from each term, multiplying by minus one and reversing the order, we get:

$$P\left(\bar{x} - 1.96\frac{s}{\sqrt{n}} < m < \bar{x} + 1.96\frac{s}{\sqrt{n}} = .95 .\right)$$

The last equation is of course the confidence interval itself. This is frequently the

first confidence interval in Statistics 200 because it is the simplest. Putting it slightly vulgar, it is lousy. Notice that we have not used any of that preceding mass of probability that makes up the first part of Statistics 200. We have used the fact the a sum of i.i.d. normal variables is normal. We have used the fact that the variance of a sum of independent random variables is the sum of the variances. We have used that the sample means of a normal distribution are normally distributed with

By trying to teach too much, we frequently end up teaching too little.

the same mean as the underlying population and with the same variance divided by the sample size. We may view the last fact as equivalent to the prior, but to the student, it is a big step. The concept of a random variable that is itself the sum of random variables is abstract to the student. The concept of a distribution of sample means is even more abstract to the student. (Note that students will use random variables that are sums of other variables with ease; this does not mean that they can explicitly handle the concept.) We have also used the concept of the standard normal distribution and we have used the normal table. Lastly we have used equalities of the form: $P(x<c) = P(ax + d < ac + d), a<0$. This is a revelation to the student. I myself have tended to use the grisly rationale that goes more or less as follows: *Let us suppose that the mean height in the NBA is 6'10". Then if we pick a player, X at random, P(X>6'10") = .5. Now suppose that we amputate twelve inches of leg from each player. Isn't it clear that P(X-12">6'10"-12") = P(X-12">5'10") = .5?*

However we motivate and teach the confidence interval above, it entails a lot of new ideas for the student and uses little of the previously introduced ideas. Normally, this is where we lose any remaining students. The student proceeds through the rest of Statistics 200 trying to understand how to use the formulas and to survive the course. Again the typical course will include both ANOVA and linear regression. There can be no attempt by the student to understand any underlying theory or why the topics are important. The reason is simple: Statistics 200 goes too fast and has too much content. The result is that six weeks after the course is over, the vast majority of the students remember little if anything of the course content.

The Source of the Difficulty

There are many reasons for a course like statistics 200 which is required of so many students but teaches very little to most of them. However, I want to concentrate on one problem. All of us that teach in the mathematical sciences tend to forget how many concepts are actually involved in a course. With time, the substance of any course that we master seems to diminish into a nice tidy package. It seems reasonable that we can teach the material, to capable students, in very short time, and when the students start to choke on the material it is easy to conclude that the difficulty is that the students are not capable.

The Statistics Specific Solutions

Since I have used Statistics 200 as my prime exhibit of a course that includes too much and teaches too little, I will begin by discussing solutions to that specific case. An obvious solution is to split the material into two courses: Probability 200 and Statistics 200. On the other hand, what if we tried teaching the basic non-mathematical statistics course without much probability? Could we do it properly? I say yes. I think a text that would serve for such a course and which shows how to do it is *Statistics for Research* by Dowdy and Wearden [8]. The focus of a statistics without probability would be data collection, data analysis, and experimental design. However, often when we have statistics 200 as a prerequisite, it is for the probability. For example, courses in operations research generally have Statistics 200 as a prerequisite but for its probability content not its statistics content--and teachers generally find that their students have to be retaught the probability from the beginning. We could instead consider teaching probability without statistics. This would give more time for probability concepts to sink in and would provide an excellent foundation for statistics. For some reason there are very few sophomore level courses in probability.

Another solution to the statistics problem is to teach a data analysis course based upon the new computer intensive methods. These methods are

conceptually much easier than the established techniques that we generally teach. They enable the student to learn techniques and not lose sight of the underlying problem. In fact, Sir Ronald Fisher may have been thinking along these lines (see the second chapter of Box, Hunter, and Hunter [9] which is, I am sure, derived from Fisher). Techniques such as ANOVA were developed partly because computer intensive methods were not feasible. Teaching an elementary course based on these techniques is the subject of [10] and [11]. A good readable introduction to these methods is given by Noreen [12].

A Last Irrelevant Remark on Statistics Education

I can't help but add one comment on statistics education. It is generally the case that when we learn statistics at a deeper level than Statistics 200 that we study *mathematical statistics*. Here our methods are calculus based and we spend much of our time doing and studying analysis. It is quite common for courses at the first year graduate level to be measure theoretic. What is remarkable to all this is that when it comes to doing statistics and doing statistical design, most of that is irrelevant. Furthermore, it is not necessary even for an appreciation of theory and fundamentals. The just mentioned book by Box, Hunter, and Hunter [9] is superb in substance and theory and does not rely on calculus. The book by Snedecor and Cochran [13] is a magnificent volume of statistical methods and makes no use of calculus. It must be admitted that both books do require substantial mathematical maturity.

The General Solution

I have stated the general problem throughout this essay: *By trying to teach too much, we frequently end up teaching too little*. The solution is obvious: sometimes at least, *Less is More*. We can obviously teach too little, and we can obviously fail to challenge the students, but if we go beyond their capacity to keep up, their tendency is to stop learning at all, and concentrate on just passing the tests and surviving the course. What little they do learn goes into short term memory and is lost. How do we decide how much material is appropriate? I know no easy solution to that problem. However, there are far too many teachers in the mathematical sciences who do not recognize the problem.

1. Cargal, J.M. "On Teaching in the Mathematical Sciences," in the *Humanistic Mathematics NetworkNewsletter #6. May 1991. 86-89.*

2. Hamming, *R.W. The Art of Probability For Scientists and Engineers*. 1991. Addison-Wesseley.

3. Conall Ryan. *House of Cards*. 1989. Knopf.

4. Vos Savant, M. "Ask Marilyn" in *Parade* . September 9, 1990.

5. Engel, Eduardo and Achilles Venetoulias. "Monty Hall's Probability Puzzle" in *Chance* Vol 4. No. 2. Spring 1991. 6-9.

6. Cargal, James M. Letter in *Recreational and Educational Computing* Vol 6, No. 3. May 1991. 6-7.

7. Tanur, Judith M, et al. *Statistics: A Guide to the Unknown*. 1989. Wadsworth.

8. Dowdy, Shirley and Stanley Wearden. *Statistics forResearch*. 1983. Wiley.

9. Box, George E. P., W. G. Hunter, and J. S. Hunter. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. 1978. Wiley.

10. Simon, Julian L. and Peter Bruce. "Resampling: A Tool for Everyday Statistical Research" in *Chance*. Vol 4, No 1. Winter 1991. 22-32.

11. Peterson, Ivars. "Pick a Sample" in *Chance*. Vol 140, July 27, 1991. 56-58.

12. Noreen, Eric W. *Computer Intensive Methods for Testing Hypotheses: An Introduction*. 1989. Wiley.

13. Snedecor, George W. and William G. Cochran. *StatisticalMethods, 7th ed.* 1980. Iowa State University Press.